DOCUMENT RESUME

ED 359 174 SP 034 615

AUTHOR Clark, Dave
TITLE Teacher Evaluation: A Review of the Literature with Implications for Educators.
PUB DATE 93
NOTE 41p.; Seminar paper, California State University at Long Beach.
PUB TYPE Dissertations/Theses - Undetermined (040)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Change Strategies; *Educational Change; Elementary School Teachers; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; *Evaluation Research; Evaluators; Futures (of Society); Literature Reviews; School Districts; Secondary School Teachers; *Teacher Effectiveness; *Teacher Evaluation; Teacher Improvement

ABSTRACT

In the spirit of education reform, this paper presents a literature review that provides a framework for discussing methods of teacher evaluation, their effectiveness in assessing what they purport to do, and the concern for better evaluation methods that can lead to improved teaching. The chief concerns addressed are what actually defines an effective teacher and how evaluators can accurately assess whether the teacher is meeting the criteria. Following a definition of terms, the following aspects of research outcomes are outlined under the headings: (1) A History of the Topic; (2) What Is an Effective Teacher?; (3) Teacher Evaluation (purposes); (4) Methods of Evaluation; and (5) What To Do with Evaluation Results. The paper concludes with the following recommendations: before states and school districts evaluate teachers they must evaluate the process at hand; districts and states must take a step back and assess their methods for evaluating teachers; and since evaluation methods as used by many states and districts have not been changed in the last 20 years, the suggestions and research outcomes as discussed in this paper should allow states and districts to move teacher evaluation into the 21st Century. (Contains 27 references.) (LL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED359174

TEACHER EVALUATION

A REVIEW OF THE LITERATURE WITH IMPLICATIONS FOR EDUCATORS

BY DAVE CLARK

SUBMITTED TO

DR. LASS

EDEL 695

SEMINAR IN ELEMENTARY EDUCATION

CALIFORNIA STATE UNIVERSITY AT LONG BEACH

SPRING 1993

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. R. Roberts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

SP834615

BEST COPY AVAILABLE

INTRODUCTION

The climate in the United States for the past 10-15 years, with regards to the perceived failure of education, is to lay blame on the educators, or more specifically, the teachers in the classrooms of our schools. This is found in the proverbial 1983 release "A Nation at Risk: The Imperative for Educational Reform" and in any major newspaper or periodical on the news stands today, let alone the discussions held at the dinner table in many homes across America. I use the words perceived failure, as there exists much debate in the education and political fields as to whether or not our schools are actually failing (and if they are, how and to what extent, and then who to blame?). Changes in the family structure, shifts in morality, the influence of the media, drugs and alcohol abuse, governmental priorities with regards to financing, and the change in the ethnic makeup of the country are all agreeable factors. Yet, the blame still falls at the feet of the teachers and the need to improve the quality of our schools. I will not belabor the point that schools can improve, but the concern of this author is the methods used to accomplish this task. The current school reform movements are targeting several diverse but integrated areas. These include, but are not limited to: revamping curriculum designs, the instructional models or approaches used in classrooms, a shift towards school-based management or local control versus the traditional perspective of a central administration, the role of the surrounding community, calender schedules (year around versus traditional, and more and/or longer days spent in school), and the evaluation of teachers.

The area of evaluating teachers or the philosophy of accountability within education, offers the belief that there are too many teachers who are not effective in doing the job they were hired to perform. The are many reasons for this perception, and some are no doubt true. Lack of proper and effective training is accepted by the general public and even by members of the profession. This may occur either through poor or ineffective new teacher training programs at the college level, or once hired as a teacher, by the lack of continued professional growth through effective staff development/in-services at the school site or district, and the need for continuing their

education at the college level (in masters or doctorate programs). In addition to these two, another cause is the idea of certificated teachers who lack the "gift" to be a successful teacher in the classroom, were hired through a fault in the system (see Caldwell, 1992 "Hiring Excellent Teachers: Current Interviewing Theories, Techniques, and Practices"), and then due to tenure, can not be fired.

The second concern, is of prime importance, as we need to ensure that the "best" educators are being hired. It will also be assumed that the training of teachers is also improving and that current movements in education are being instilled in new teachers at the college level. The emphasis of this paper will therefore focus on methods of teacher evaluations, their effectiveness in assessing what they purport to do, and the concern for better evaluation methods that can lead to improved teaching.

STATEMENT OF PROBLEM

If the current school reforms are to be implemented, the need for qualified and effective teachers will be much greater. For reforms to take place, the most effective and knowledgeable educators must be in the classrooms seeing that the recommendations are being implemented as outlined in the reform proposals. The need to insure that the evaluation methods used for assessing a teacher's performance are measuring what is assumed is paramount. Within the educational field, there exists the general perception that the current methods of teacher evaluation, can and do measure what they purport. This perception though, is not shared by all. There are those teachers, educators, and members of the public who feel that the current methods used to evaluate teachers may be missing the mark. The chief concerns appear to be what actually defines an effective teacher, and then how can you accurately assess if the teacher is meeting this definition? On the same level, what if the teacher is below the expectations, is the teacher to be terminated or put on probation? What if the teacher is exceeding the expectations, should there be a means of rewarding those who can show a level of superiority, and then therefore a method of punishing those who are found to be lacking?

These are sensitive issues, with unions and labor concerns all deeply involved. Granted the methods used to evaluate teachers must safeguard against unfair labor practices or vindictiveness and favoritism by school administrators, and many of the current evaluation methods do just that. But certain inconsistencies exist when teachers are evaluated. Many teachers maintain the status-quo with little signs of creativity, academic or professional growth, or gains in student achievement. Yet these teachers continue to receive evaluations that seem to say "performing at an acceptable level of professionalism", and those teachers who are doing the great things in the classroom or with their students, end up with a similar level of evaluation with little or no reward for their achievement. Is the problem with the criteria that defines what is an acceptable teacher, or are there too many variables, such as bias or interpretations by evaluatees, that create this disparity?

This author would have to concede that both are to blame, though the former may be of greater concern. Too many districts or schools have allowed the profession of teaching to lower itself to a common denominator when evaluations are being defined. The movement has been, for too long, of a gradual shift towards mediocrity in our society, and as a reflection of such, in our schools as well. While this may cause a few cries of anger from the teaching profession, the vast majority of teachers must admit, that while we try our best to educate others and ourselves, way too frequently we see teachers accepting their role as a member of the status-quo. The role of teachers must be to elevate not only what we expect of our students, but what we expect of ourselves. The problem lies in what truly defines an effective teacher, not in the definition of one who meets the minimum standard criteria. Therefore this paper will also address this definition as well.

DEFINITION OF TERMS

While the following terms may have obvious meanings for the reader, they will be used in the text under the following definitions.

Evaluation: Any means that is used by school site or district personnel to assess the evaluatee in terms that define the district's or state's education or classroom teacher requirements. Several specific evaluation forms and methods will be discussed, with the paper defining the assumed "ideal " form to use in evaluating teachers. The term evaluation may be substituted with the terms assess or assessment with no change in the stated definition.

Summative: With respect to teacher evaluation, this implies an evaluation method that is based on the summation of all methods used in the assessment of a teacher's performance. An evaluation may be summative and not necessarily formative, though both may occur in the process.

Formative: With respect to teacher evaluation, this implies an evaluation method that is based on the formal meetings between the evaluator and the evaluatee to determine the teachers performance. These meetings may occur as observations of the act or discussions after the fact that are intended to increase the degree of learning in the classroom through the sharing of information An evaluation may be formative and not necessarily summative, though both may occur in the process.

Evaluator: Any person or persons who are involved in evaluating the targeted teacher. This may include teachers or administrators working within the educational field.

Teacher: Anyone working as an educator in the classroom setting with students, either publicly or privately, who is evaluated using the criteria as stated under the terms of their school, district, or state of employment.

Instruction: This involves the transfer of knowledge and skills, as defined by any state, district, or school curriculum. No attempt is made to define what is an effective instruction, nor is the definition limited to the degree of this transference. The term teaching is interchangeable with instruction for this definition.

HISTORY OF TOPIC

The evaluation of teachers grew out of the community's need to determine job continuation (does the teacher stay employed by the local governing board), to set pay increases, and for promotion. The area of pay was usually argued for, with good evaluations used as a means of establishing credibility by the teacher. The former though, usually was used as a means to eliminate teachers who were either poor, unpopular, or controversial. This was also used as a means to justify why teacher A got the advancement but not teacher B, though no doubt, nepotism and/or favoritism played an active role. These evaluations were loosely based on what the local community or school board felt best reflected their needs, but not necessarily those of their students. While educational guidelines existed, in that curriculums were defined with desired outcomes, the general expectations dealt more with local societal beliefs, such as: shared religious principles or doctrines (god and the family), national and civic concerns (democratic principles, love of the country, perceptions of the world), and moral responsibilities. Since the teacher was hired by the community as an entity unto themselves, in the sense that schools were small, they had little say in the matter. The setting would be the same, if not worse in the true "private" schools, since all control belonged to the governing board.

The industrial revolution brought an advancement in the teachers' power to determine their place in the educational sector. As the country changed from rural farms with widespread populations dictating small local schools, to one of centralized cities with large and growing populations, the schools grew in size and stature. This change in the schools brought larger classrooms, single grades ranging from the primaries to upper levels, and a greater need for qualified teachers who could teach the new student. Over time this also brought the emergence of unions to protect the teachers from unfair practices by administrators. The vast majority of public school teachers were women, who had (and to some extent still do) little political power to protect their professional standing. What the unions attempted to do, was to set specific criteria for evaluating a teacher in terms of dismissal, continuation, or for advancement. The criteria was at

best very basic and generalized, and with the profession still occupied predominantly by women, lacking the political power to back it up. The local communities, districts, or states still determined the who and how of evaluating teachers, though educational criteria was now becoming a greater focus.

In the early 1950's more and more men started to enter the public school system as teachers. These new members brought with them their political power and male values and characteristics, causing the emergence of a new public perception of schools. Teaching was now seen as a profession that was not female oriented, though men had long dominated the private schools and specifically the higher levels of education, but one where the male was equally accepted. Time also influenced this perception with the emergence of the cold war between the United States and the Soviet Union. This placed a greater emphasis on the "product" being produced by the school system. The concern being that the Soviet Union was producing better educated students who would then lead their country to a cold war victory.

This caused a new push to find better and better teachers within our country. The search lead to a greater number of male college students being "called" to the profession. As more men entered the profession, unions grew stronger through their new affiliations with the great labor unions of the blue collar workers. Their influence and role in the evaluation of teachers offered the profession the respect long over due. Perceptions of the school system were positive as students emerged ready for college, industries and democracy prospered, all resulting in America enjoying a sense of educational security.

The shoe fell in 1983 with the release of "A Nation at Risk: The Imperative for Educational Reform" by the National Commission on Excellence in Education. This document revealed that the schools were not producing the great students we had assumed they were, but in actuality, were failing to teach the basic skills of reading, let alone the rest of the curriculum. Education, over the past 30 years, had evolved into system based on the premise that teacher-proof curriculums, test-based instructional management, and student competence testing alone would improve learning. These policies assumed that adherence to a predetermined teaching format would result in the

desired level of learning. Teachers were viewed more as laborers implementing a prescribed program in a manner determined by policy makers further up the educational hierarchy, than as professionals with a repertoire of techniques and the ability to decide for themselves how techniques should be applied. This perception of teachers led to the enactment of rigid regulations and mandates designed to improve the quality of education without actually addressing the competency of those hired to "process" students through the schedules, curriculums, and exit tests required by the schools. Policy makers sought to "fix" the problems in education by enacting more regulations. Student failure to achieve higher level learning was attributed to the nonconformity of the schools and/or teachers to the prescribed methods of education. The solution to this problem was thought to be more detailed curriculum prescriptions and more careful monitoring of their implementation (Alexanderov, 1989).

What the 1983 "A Nation at Risk: The Imperative for Educational Reform" by the National Commission on Excellence in Education called for was a movement away from the cookie-cutter approach to teaching and towards the emergence of effective teachers who would lead our schools to the new levels of excellence. This would occur through using the established methods of evaluation, as based on the descriptions of public education by Alexanderov. These evaluations were not designed to identify effective teachers, but those who met the minimum requirements to be a teacher. The requirements included punctuality, providing a safe learning environment, and upholding school rules and district policy (Alexanderov, 1989). Unfortunately, some of these are still present in the evaluations used today.

What is so disturbing with the statement by Alexanderov, with respect to the perceptions of teachers and the dire situation within the public schools, as well as, the findings from the National Commission on Excellence in Education, is that the National Education Association (NEA) had already outlined what was wrong with the public schools and the teachers' misplaced role in its function. This appeared in the book "Schools For The 70's And Beyond: A Call To Action" (1971). This book clearly outlines the faults within the education system; the factory approach to educating students, failure to recognize the individual student needs, teachers as laborers - not as

professionals, a centralized administration that lacks any sensitivity to the actual educational process, and a curriculum that offered little relevance to the parties involved (teachers and students alike). This book, while written over 20 years ago, offered an alternative to the then present educational system that today would appear quite topical. In brief, the book saw the teacher as a true professional who should have greater control in the classroom with respect to the textbooks, instructional strategies, and curriculum designs with a more humanistic attitude towards the child in the classroom. In addition, the report proposed the decentralization of administrative power, returning it to the school site or local community for greater school effectiveness in educating students. This book also offered a definition of what the NEA felt would be an effective teacher and how the evaluation of teachers would need to have a role in the emergence of a new professional growth within the ranks of all educators. This early definition of an effective teacher, mirrors those of the research today.

Many districts have since developed an educational program that reflects these recommendations, both in terms of curriculum designs and the teacher's role in the educational process. Unfortunately, time has not been nice to the public school system. As the introduction stated, our school system is under attack again, and districts have attempted to remedy the situation by the methods used in the evaluation of their teachers. The problem lies in what the districts are really trying to evaluate and the how these evaluations are used in the educational system. To address this, we will need to look at three areas of concern, what is an effective teacher, how can they best be evaluated, and what can we do with this evaluation?

WHAT IS AN EFFECTIVE TEACHER?

The major problem with defining an effective teacher is deciding whose definition to use. In addition, there is the need to keep a distance from the term competent or competency. I find this term to be a serious fault in the system, as too frequently it has come to mean "meeting the

minimum requirements." This term should only be used in testing students for admission to a credential program for teaching and then into the actual profession. The problem lies in teachers who accept this as a measure of ability and feel that their doing fine. This is so wrong and must be rectified through change in the perspective of the teacher's role in the educational system.

In choosing the definition for an effective teacher, the research in the educational field offers us many choices. Obviously, the definition involves someone who can increase a student's knowledge, but it goes beyond this in defining an effective teacher. Vogt (1984) uses the findings of a task force for the Campbell County School District in Gillete, Wyoming called the CET or "Criteria of Effective Teaching." Vogt cites the CET in saying that an effective teacher is one who provides for different student abilities, assures that students were aware of expectations, incorporates different instructional strategies into the lesson, involves the student in the learning process, uses instructional objectives outlined in the curriculum guides, uses written outlines or course outlines, assesses academic and social needs of students, assesses academic and social growth of students, assesses effective learning modes of students, understands the scope and sequence of the course, and can plan with flexibility. While this sounds monumental, there are teachers who do this and more every day.

The CET continues with four individual areas that further define an effective teacher These are, (1) in classroom management, by providing a social-behavior environment that enhances learning, establishes effective discipline practices, and maintaines a physical environment in the classroom conducive to learning, (2) having good student relationships, by involving students in decision-making, helping students develop a positive self-image, dealing objectively with conflicts between students, and aiding students to develop social-interaction skills, (3) in their professional qualifications, by maintaining positive relationships with all employees, establishing effective lines of communication with parents, citizens, and other adults, and handling general school responsibilities, and (4) through their personal demeanor, by demonstrating an ability to understand, appreciate, and respect young people, demonstrating maturity, and communicating effectively in oral and written form.

To be added to the definition by the CET, is the need as suggested by Ernst (1982), to evaluate teachers on their ability to design and implement instruction utilizing educational technology. The emergence of technology in the classroom (computers, video systems, and multi-media components) requires teachers who are trained in their effective use in instruction to meet the new curriculums. This is important, as the integration of technology into classroom instruction offers significantly greater learning in less time and increases motivation to do so (Ernst, 1982). For this to occur, in-depth instruction and training must be in evidence from the pre-teacher education programs in colleges to school site in-services. A survey of media specialists, teacher educators, administrators, and teachers was conducted to measure the importance of the need to assess technological competencies among teachers. Ernst found that on 69 identified competencies, 64 were rated as "moderately important" or "very important." Ernst summarizes that the study identified 69 educational competencies that are perceived to be related to teaching effectiveness. An assessment of a selected sample of teacher education programs revealed a significant percentage of these competencies are not included in the programs. This disparity offers a challenge to personnel in teacher education programs who seek to improve teacher effectiveness. In addition, it may be perceived as a challenge to improve teachers who already in the classroom through the use of in-services offered by the school districts.

Collins (1990), in a discussion on the TAP (Teacher Assessment Project by the National Board for Professional Teaching Standards), defines an effective teacher through a summative approach. This definition mirrors that of the CET, by stating an effective teacher is someone who is (1) committed to students and their learning, (2) knows their subject matter and how to teach those subjects to students, (3) is responsible for managing and monitoring student learning, (4) can think systematically about their practice and learn from experience, and (5) is a member of the learning communities. These are all fine, except that they are open too much to interpretation by the evaluator and the evaluatees. It is important that a definition is clear and concise for all parties involved. To further this controversy TAP takes the position that there is no one way to teach, each being valid in its surroundings. Within the definition offered above, TAP recognizes the

individual and the existence of other forms of teaching that may be novel or on the cusp of current trends. TAP also suggests that teaching takes place in a unique context, at a particular time, and with specific students. The criteria used for one class may not relate to another class based on differences among students, grade levels, time of year, and geographic location. The implications for assessment will be discussed later, but one can see a need for flexibility when defining an effective teacher, let alone when evaluating a teacher's performance.

Swank, Taylor, Brady, and Freiberg (1989) offer a much more narrow definition. The authors see an effective teacher as one who (1) increases academic questions to individual students, (2) decreases the extent of academic lectures, and (3) decreases nonacademic variables such as negative feedback, simple questions, dictating information, providing corrections, and rudimentary reinforcement. The authors felt that these qualities were more easily identifiable when assessing a teacher's performance, than those dealing with personal attributes.

Million (1989) offers a multiple-strategies approach to evaluating teachers by identifying ten areas deemed necessary for effective instruction. These are (1) the classroom climate in its totality (the summation of a consciously created environment and its often unintentional effective and physical dimensions), (2) the opening lesson or simply the opener the teacher uses to create and then maintain student interest, (3) the use of instructional objectives in terms of the desired cognitive goals, (4) justification of the content to create a purpose or meaning for the students, (5) selection of content being taught, (6) the teaching strategies used to meet the needs of the students, (7) the review of material to ensure learning, (8) lesson evaluation in terms meeting the desired objectives, (9) lesson evaluation in terms of assessing student achievement and/or progress, and (10) classroom management that focuses on student discipline to maintain the needed environment.

The definitions of Collins, Million, and Swank, et al. all point at a highly trained and gifted individual, who possess a functional knowledge of the curriculums to be taught, and knows how to effectively teach them to divergent student populations while meeting the needs of the individual.

One area that was avoided by most authors was the idea of using student achievement as a measure of effectiveness. It is easily argued that an effective teacher will have students learning,

that is, their achievement will increase. After all, if a teacher is doing an effective job at teaching (say the teacher fits the previous definitions), one would expect the students to learn more than a teacher who is perceived as not being as effective (one who does not fit the previous definitions).

The problem is in determining how best to measure this student achievement. Teacher observation, portfolios, peer assessment, and standardized tests are by far the more common. While their use is recognized as a valid measure of student performance, they suffer from variability among the teachers involved and may not be reliable. Only the standardized test is used in large enough numbers to offer comparable measures of teacher effectiveness. This test has many different forms, depending on the state or district under consideration (CAP, CTBS, ITBS, etc.). These current measures of student learning, standardized tests, are simply too easily influenced by extraneous factors to make them a valid and reliable measure of teacher effectiveness. The key is reliable, as these measures may offer an alternative perspective when evaluating the performance of a teacher. The authors Bingman, Heywood, and White (1991) suggest such an alternative, and will receive greater attention when we discuss the methods of assessing a teacher's performance in the classroom.

We appear to be able to identify what an effective teacher should be, and we can test or show for student achievement, but do they really relate to each other, and if so, then how? White, Wyne, Stuck, and Coop (1987) suggest from their research that there is no single teaching behavior that is strongly related to student achievement, but that a cluster of teaching behaviors occurring together can reliably distinguish effective from less effective teaching in most settings. The model as presented by the authors, is seen as one that will lead to an increase in student achievement. This was determined through the authors' field testing and statistical research reviews. The authors present five teaching functions: (1) management of instructional time, (2) management of student behavior, (3) instructional presentation, (4) instructional monitoring, and (5) instructional feedback. These are subdivided into finer behaviors to be exhibited by the teacher in their Taxonomy of Items in the Teaching Performance Observation Instrument. This offers a clear picture of an effective teacher, while still allowing for freedom of personalities and the

individual based on the interpretations of each item. The authors state that their research could not define all that a teachers does to cause student achievement, as the role of teaching is too complex and personal. The assumption is that no one set of specific teaching practices can possibly account for the totality of the teaching process under all conditions. It should be noted that the authors fail to define student achievement in their research, which causes concern as to the validity of their model.

White, Wyne, Stuck, and Coop (1987), relate that teaching must look to existing empirical research to identify those teaching practices that are consistently related to student achievement. While the authors have attempted to create a model based on their own research, the authors see neither the general practice of teaching nor the training of teachers, as being systematically guided by existing empirical knowledge about the relationship of various teaching practices and student outcomes. Rather, the personal wisdom and the consensus judgement of educational authorities have defined the appropriate teaching practices as discused previously. There is nothing wrong with using personal wisdom developed through classroom experience and the opinions of recognized educational leaders to identify what works and what doesn't when teaching. The problem is that, what one teacher finds to be effective, another teacher may see it as being ineffective.

This brings us back to the position taken by TAP, in that there is no one way to teach, and considerations as to context, time, and geographic locations are all factors one must consider in assessment. The definition of an effective teacher must allow for diverse approaches to instruction, but this must occur within the desired model of an effective teacher. Districts and schools will need to be flexible in allowing for new approaches that offer an alternative to the tried and true approaches to instruction. If these new approaches produce achievement in students and can fit within the definition of an effective teacher, their acceptance must be forth coming. In addition, unless there is a consensus among all educators (evaluators and evaluatees) as to the definition of an effective teacher, then the process of teacher assessment would become chaotic.

The descriptions offered so far, all tend to mirror each other, while either adding or extrapolating some item. Therefore a general consensus does exist in defining an effective teacher.

From the previous definitions we can assume an effective teacher should fit the following summation:

- Knows and understands the stated curriculum's content relevant to the specific grade level, its relevancy to the students, its purpose, and its goals.
- Offers the student total instruction through a variety of approaches that reflect current research to meet the stated curriculum.

  This would include planning, presentation, the instructional strategies employed (including technology), integration with other contents, assessment, and feedback.
- Offers a total environment in terms of the physical, personal, and instructional parameters.to meet the needs of the individual and the collective.

  This would include the look and feel of the classroom, and the need to fulfil certain intrinsic needs in the student(s).
- Increases student knowledge and achievement.

  As measured on achievement tests or other selected measures.
- Maintains a professional demeanor with students, parents, and educational peers.

  This would include appearance, attitudes toward the profession and others, and the need for professional growth.

TEACHER EVALUATION

PURPOSE

Teachers are evaluated for several purposes. While most of these have been dealt with in the historical discussion, concerns with competency, professionalism, advancement, and merit pay are primary issues. In addition, student achievement must also be included. Obviously the goal of teaching is to cause an increase in knowledge (either measurably or immeasurably), and teacher evaluation may be the prime factor in this achievement. Not so much by seeing if a teacher is teaching, but by seeing how a teacher can become a better teacher. Teacher evaluation should be a driving force in selecting staff development topics. This can then be used to address identified shortfalls or weaknesses in the staff. Newton and Braithwaite (1988) found that teachers saw little actual purpose to evaluations, though the teachers' own perceptions placed a high value on evaluations. Evaluations were simply a means to an end, with little impact on their day to day existence within the classroom. Teachers felt that evaluations, while assessing their abilities, must lead to feedback and improvement in their own profession. This was seen to be seriously lacking under the current system.

This author, as well as Buttram and Wilson (1987), suggests that evaluations may best be used to identify the more effective approaches used in teaching, and using this knowledge to drive staff development and possibly teacher training at the college level. The current reform movement in education will require a reform in the purpose of teacher evaluations. While the need to determine levels of competency, professional and/or pay advancements, and tenure are all well and good, the need to adopt effective teaching approaches and curriculum models is of greater importance.

METHODS OF EVALUATION

Davey (1991) offers an excellent introduction to assessing the performance of an individual. Davey sees the concept of performance assessment, as an employee testing strategy, in having a long history. This particular history is in regards to the trades and labor jobs, where apprentice blacksmiths, carpenters, or painters must prove the mastery of their craft by performance. In these cases, "scoring" might involve simply judging the acceptability of the product, which can be seen, felt, examined, and therefore in some way compared to a standard. However, the assessment task becomes more difficult when the primary outputs by the candidate are not concrete products but processes - decisions, actions, interactions, explanations, and so on, that vary from candidate to candidate and have no single objective standard to use as a scoring template. In more recent times, a process-oriented example of a performance assessment model has emerged in the form of the assessment center.

This center or process as Davey sees it, is composed of the following components: (1) the conduct of a job analysis to identify the dimensions of effective job performance (these dimensions, or a subset of them, then become the focus of the assessment), (2) the use of multiple exercises (multiple observations representing a variety of situations), (3) the use of multiple assessors (diversity of age, ethnicity, gender, and perspective to eliminate bias), (4) systematic procedures to enable the accurate observation and recording of behavioral observations (which requires a thorough analysis and codification of the potential strategies and responses of the candidates), and (5) thorough assessor training.

The question is then, how does this relate to assessing teachers? The most obvious is to take the model as outlined by Davey, base it on the definition of an effective teacher, and set standards for teachers to meet. This is exactly what the majority of districts have done, but with poor results. Blecke (1982), describes how the procedures may consist of a request for lesson plans, a visit to the classroom using an archaic checklist instrument that offers a simplistic view of teaching effectively, and a follow-up conference with the teacher. Unfortunately, the assessment

of teachers may be too complex for such a simple approach, with other factors causing poor assessments (such as those suggested by the TAP). Davey even admits that teacher competence involves a complex set of knowledges, abilities, and personal attributes all involved in a dynamic interplay within the classroom environment.

The need is for a holistic approach, in which a far more complex and complete performance on the part of the teacher is elicited and then evaluated. Pembroke and Goedert (1982) suggest teacher evaluation models must be accepted as fair and objective by teachers, be related to the specific requirements of the job and the unique needs of the profession, specify the factors against which the teacher will be measured, reliably measure teacher performance and specify by whom and how the measurement will be done, clearly communicate the expectations for performance to the individuals, and provide for teacher development as a part of the process. Pembroke and Goedert also stress that teachers must be involved in the whole evaluation process for it to have any real value to the educational system.

Dunkleberger (1982) suggests that principals look for the following when visiting a classroom; evidence of planning, setting realistic goals, how the materials are prepared, the use of instructional aids, the setting and pacing of the lesson, evaluation and feedback to the student, and closure. Dunkleberger also sees the need for evidence of student motivation, quality and a variety of activities, communication, and good classroom management. While Dunkleberger offers a clear definition of what is desired in a classroom (of which some do not imply effective teaching or student achievement), the author provides no method of assessment. The fault with Dunkleberger's suggestion, is that this is the current method of evaluating teachers used by various administrators. If the items as described by Dunkleberger are evident in the classroom, than the assumption is made that the teacher is doing their job. The implication is that if it looks good, than it is good. Unfortunately, it is just not that simple.

Redinger (1988) suggests that there is no one way to assess teachers, with each district and/or school needing to develop a method that best fits their needs. Most methods of evaluation have inaccuracies and disadvantages when used alone or without sufficient communication with the

teacher and evaluator. Successful systems have been shown by research to have several attributes in common. The evaluation plan matches the goals of the school district, and the district commits sufficient time and resources to the chosen plan. The chosen criteria in successful plans are based on the research on effective teaching methods, and the use of a multi-source data program gives a total view of the teacher's performance. Redinger sites research that shows formative evaluations helping teachers to change their performance by setting goals or developing a plan for more effective teaching. After a cycle of formative evaluation, summative evaluation is used to give a final rating of effectiveness or for personnel decisions.

The following are some suggested models as based on the previous definitions of an effective teacher and the needs as delineated by Pembroke and Goedert. The intent of this survey was not to review all existing evaluation methods already in use (the California STULL, the Tenenessee Carrer Ladder, or the Program for Effective Teaching as based on Madeline Hunter's ideals to name a few), but instead various models based on the research that offer a means of improving teacher assessment.

The CET, as reported by Vogt (1984), uses an evaluation tool that assesses a teacher's performance on the criteria as defined by the CET for an effective teacher. This criteria has already been outlined previously in this text. The evaluation consists of classroom observations and summative evaluations by school administrators. The evaluators received training in how to use the evaluation and observation tools. Four ratings of performance are available in each category of observed behaviors in the classroom. These are: (1) Exceeds District Expectations, meaning the teacher's performance is well above the norm, (2) Meets Districts Expectations, meaning the average teacher, (3) Needs Improvement, where a teacher falls below district expectations within a specific area, and (4) Unsatisfactory, where a teacher falls below district expectations within a several areas.

Million (1987) suggests a multiple strategies model to assess a teacher's performance. This model reflects Million's definition of an effective teacher as previously discussed, but presents a slightly different perspective than those based on the model standard by Davey. Million sees

teachers and administrators uncomfortable with the traditional model of evaluation, in that it poses questions of validity, and causes anxiety and hard feelings based on poor evaluations. In addition, teacher behavior during assessment is not indicative of what they normally do in their classrooms. Curriculum, classroom climate, rules of behavior and other classroom dynamics often change when assessors are present, only to return to their former status following assessment. This is more common among teachers who perceive themselves as simply adequate, let alone those who lack the qualities of an effective teacher. These teachers will prepare their classrooms for the evaluation, thereby presenting a false picture of their personae as a teacher. Unfortunately, this occurs more often then one would like to believe.

The model as posed by Million, stress a communication between the evaluator and the evaluatee as to the parameters to be assessed. Both of these parties are trained in the model, creating a more effective teacher and administrator. That is, do both parties agree and understand what is to be observed and evaluated and how a teacher can be effective. This model follows closely the definition of an effective teacher as suggested by this author and the available research. The model also does not specify the curriculum or teaching strategies to be employed in the classroom, thereby allowing greater freedom for the teacher to present alternative means of instruction, while still staying within the definition of an effective teacher. Being in part a summative evaluation, administrators can conduct long term evaluations of teachers and collect copious information from which to draw meaningful conclusions.

The following model by Collins (1990) is taken from TAP (Teacher Assessment Project by the National Board for Professional Teaching Standards). This model, based on their definition of an effective teacher as previously discused, considers the complexity of the teaching profession, as did the previous model by Million (1987). Based on the premises of the diversity in teaching strategies, the context that it takes place in, and the influence in time, TAP makes the assumption that the assessment of teachers will require a battery of modes for assessments to be valid (a summative approach). Collins continues by suggesting that no one mode of assessment is going to be sufficient. TAP places an emphasis on the use of observations, simulation exercises,

portfolios, and portfolios based in simulations. The use of portfolios for assessing teachers reflects their use in the classroom for the evaluation of students. Here, a teacher's portfolio would contain documents that would provide evidence of the knowledge, skills, and dispositions of the teacher. A teacher's selected "best work", lessons that were effective in teaching, would be an integral part. Additional material would be student's work or progress representing the teacher's accomplishments, and any items that signified special recognition by peers or administrators. The idea of the portfolio is of a summative approach to evaluation, in that it would represent a period of several years. This is a very novel idea, and should receive a far amount of attention, both positive and negative. In addition TAP, sees the assessment of teachers as best being accomplished by and for teachers. This relates to the idea that what teachers do in their profession cannot be explained in simple terms. The act of teaching is too complex and integrated into a teacher's actions to be easily quantified and then systematically evaluated.

This author agrees with Collins, in how controversies will exist as to the legitimacy of a teacher determining what will compose their portfolio. This may be avoided by the use of peer assistance in creating portfolios and the needed assumption that teachers will present an honest representation of their teaching experiences. The use of video tapes to document teacher's instruction and behavior in the classroom, may offer a better record of performance.

Savage (1982) also recommends the use of a portfolio, but prefers the term "Artifacts of Teaching." These artifacts would be lesson plans, tests for students, laboratory/special project activities, bibliographies and supplemental reading lists, samples of student work, peer testimony in the form of critiques, student test results on standardized tests, and any other items that the teacher felt would best represent their abilities. These would be used in conjunction with classroom observation to help fill in the blanks.

In her paper on teacher evaluation, Alexanderov (1989) finds little support for the actual use of teacher portfolios when describing their use in the Tennessee Career Ladder Program for assessing teachers. The use of portfolios in this program proved to be an exhaustive job, with little direct improvement in classroom instruction, though teachers still felt a need for their continuation.

Reasons may relate to a fault in the formative evaluation process, where the evaluating administrator failed to recognize and validate the importance of the portfolio's contents.

The STAR (System for Teaching and Learning Assessment and Review by the state of Louisiana) is reviewed by Hill (1991). The STAR assessment model reflects the current research on effective teaching and learning. Hill describes the STAR model as a method of evaluating teachers by assessing both the teaching and the learning taking place within the classroom. This is accomplished through the observation and recording of the teacher's and the students' actions, behaviors, and responses during instruction, by school principles and master teachers who are trained on the STAR system. Documentation of important student/teacher interactions, various physical classroom and learning environment conditions and/or events are recorded. Upon completion of the observation, the assessor complies and synthesizes the observation notes and uses the STAR assessment indicator to rate the teacher. These indicators are based on classroom and behavior management, the total learning environment (physical and emotional), and the enhancement of learning by the teacher (methods, materials, skills, pace, and feedback). A teacher is rated as either "acceptable" or "unacceptable" on each indicator, but only after considering the whole picture. This involves scanning the content of the indicator, reviewing all pertinent classroom context and observational data contained in the notes, and considering/comparing various examples and considerations contained in the Annotation and Decision-making Rule for the indicator. This, Hill states, ensures that assessment decisions reflect as much as possible the holistic classroom environment and teaching/learning context. This is both summative and formative, in that the assessor works with the teacher in determining the level of performance, and the final evaluation is shared between both parties with the intent to generate greater learning in the classroom.

There are other approaches to assessing the performance of a teacher, that are considered by some as being too controversial. These may include student evaluations and measures of student achievement. Of the two, student achievement may be the most feared and controversial. Even so, this approach is gaining support as a measure of teaching in the classroom. The STAR

assessment model considers student achievement, but not to the degree as is being recommended by the following authors. Bingman, Heywood, and White (1991) suggest that if we are to really evaluate teachers, then we must move away from subjective evaluation for teachers by principles, peers, and even students, and take into account the many factors over which teachers have control. The authors admit that there are many factors beyond teaching that can influence student performance. The variables of individual student characteristics (sex, aptitudes, attendance, early childhood experiences), family background (size, parental education, occupations, income, expectations), schools (expenditures, environment, philosophies, services, class size, grouping), and peer group input (social class expectations, ability) all are important factors that teachers have little influence over in determining a student's learning. These are no doubt, not all of the variables that may influence a student's learning, but may best represent the primary factors involved.

With all of these factors being influential over a student's learning, how can one determine the effects of a teacher's role in the educational process? Bingman, Heywood, and White suggest that the measurement of the influence of teaching can be viewed as a residual. If nonteaching influences explain 80% of the variance in student performance, then teaching might explain some portion of the unexplained variance. The authors did this by identifying a series of variables that were believed to be related to student achievement (see above). Next, to explain as much of the variance in these scores as possible, they entered these variables into a multiple regression equation with standardized tests scores as the dependant variable. Then, the authors used the resulting regression equation to predict scores for each child. The predicted scores were then subtracted from the actual scores for each child to produce the residual. The residuals were averaged by school to see if certain schools seemed to raise student scores above the predicted. The residuals were then aggregated by classrooms to see if certain teachers were able to raise or lower student test scores (on the Iowa Test of Basic Skills) significantly above or below the predicted.

Results supported the authors' assumptions suggesting that teachers can be evaluated using this method of predicting student performance. In addition, effective and ineffective teachers are easily discernable, as well as schools. Bingman, Heywood, and White admit that this approach

does not attempt to identify what a teacher is actually doing in the classroom. The authors state that, "We can identify a ranking of schools and teachers based only on relative student performance. We cannot identify either what makes particular teachers successful or even any of the in-class characteristics of success beyond high levels of test performance. Moreover, we do not attempt to isolate the characteristics of teachers and schools that correlate with high teat scores." The authors see the "totality of teaching" as being more important than identifying the nuances or particulars of an effective teacher. It would appear though, that follow-ups as to what these teachers and schools are doing, would help to identify what the effective methods or strategies are in teaching their students.

While limited by population factors and the focus on the fifth grade as its targeted group, this approach strongly offers an alternative method for assessing a teacher's performance. Factors of cheating by student on the tests or coaching by teachers to increase student scores, are all valid concerns. While these two concerns may already exist, close supervision and monitoring of the actual testing may help to eliminate their influences. The normal response to this, is that if teachers and students are doing their job, cheating and coaching would become superfluous to the process. The need will be to use this approach as an addition to the observational methods, and not as a substitute. This approach may also offer a means of determining merit increases as well, though if the rewards are too attractive, cheating may become a reality.

Capie (1986) in a study on the use of the TPAI (Teacher Performance Assessment Instrument), finds support for the use of this alternative assessment instrument for comparing teacher effectiveness and student achievement. The TPAI is an instrument that consists of eight teaching competencies, each of which is defined by three or four indicator statements. The indicator statements are each defined by four descriptors. Capie found that the TPAI is a valid and reliable summative measurement of a teacher's effectiveness in producing student achievement. This instrument, while intended to assess teachers for accreditation, may be used as a means of assessing a teacher's performance, with the added ability to predict student outcomes. Further studies will need to be done with experienced teachers, to see if this is a valid assumption.

Redfield and Craig (1987) also consider student achievement as a method of determining teacher performance in their report on the "Student Achievement Project" for the state of Kentucky. This project dealt with the desire to include student achievement, as based at an expected level, as a defensible measure of teacher effectiveness. Student achievement was not based on the use of standardized tests, but instead the identification of specific goals for both the student and the teachers involved. These goals related to learning and teaching outcomes based on district and state curriculums and were agreed upon through conferences between the teachers being evaluated and their administrators. The authors see a potential for the inclusion of student achievement based on their project, but not on the use of standardized tests. The results of the project suggested that the piloted procedures, as described in Redfield and Craig research, have a potential for development as part of a teacher evaluation system that includes student achievement outcome data. When assessing a teacher's performance, this author would prefer to see the use of such a model, in preference to the use of standardized scores, for reasons of accountability and ease of use. The models, as suggested by Bingman, Heywood, and White and Capie, may also be of interest, though their use may require greater commitment and investments by the schools or districts.

Another alternative is the use of student evaluations. John Savage (1982), suggests that using the means of student evaluations may offer a reliable and valid assessment of a teacher. Savage sees the evaluation form needing to address the areas of stimulation of interest, clarity, knowledge of subject material, fairness, preparation, enthusiasm, friendliness, helpfulness, and openness to other opinions. Though a different Savage (1986), found in a study of teacher assessments that were then combined with student evaluations, that this data had little influence on a principal's final judgement. While Savage felt that a student's opinion offered a degree of validity based on the research evidence, the principals saw this as being less credible. This may be attributable to the age of the student who is evaluating the teacher. The recommendations of kindergarten through third or fourth grade students may not be a valid measure of a teacher's performance, due to such attributes as maturity and knowledge of the instructional curriculum.

Students who are older and more mature may be a more valid and reliable assessor of teachers for these reasons, though their lack of innocence may color their opinions.

Newton and Braithwaite (1988), on the same topic, suggest that by virtue of the time individual students spend in the classroom, they probably have more evidence on which to base evaluation judgements than any evaluator. Even young students can, and often do, express insightful opinions about curriculum content, methodologies and teacher attitudes. Students are also more numerous than any other group of evaluators, and therefore collectively represent a significant body of opinion. Newton and Braithwaite suggest that in terms of improving instructional strategies, student recommendations may have considerable merit. This author would have to agree, as they are often the best critics. In reference to the study by Savage, even though he found little support for student evaluations, the author felt that the perceived quality or quantity of observations by the principles may have been of a greater influence in determining their final assessment. Student evaluations can offer an addition to the evaluation process with respect to the desired outcomes - student achievement. Obviously, one would not want to base the evaluation of a teacher on student evaluations alone. Redinger (1988) suggests that student evaluations be used as supplemental material for formative evaluations, and not as the only source of assessment scores.

The quality or quantity of observations introduces the topic of classroom visitation for the observation or assessment of a teacher's performance. This author and Million (1987) suggest that teacher behavior during assessment is not indicative of what they normally do in their classrooms when observations are pre-arranged. The teacher is aware of the date for their observation, and can therefore prepare their classroom, lessons, and students to create the impression of something better than what really exists. This, it is hoped, is countered by an administrator who is aware of the "real conditions" through frequent visits to the classrooms. Unfortunately, many labor agreements limit the use of frequent visitations to classrooms in assessing a teacher's performance. These labor agreements stipulate that assessment must be based on the use of pre-arranged formal observations.

Therefore, the observations must be done effectively for their use to be valid and reliable. This will be necessary, regardless of the model being used in teacher evaluations. Swank, et al. (1989), suggest that observation be divided into two measures: micro and macro. The idea of macro measures involves assessing the effectiveness of the teacher in teacher-to-group situations by use of the Stallings Observation Instrument. This is a paper and pencil measure, by direct observation (through five 5-minute interactions or an FMI that are systematically dispersed over the whole class period, and five 1-minute interactions that are random - referred to as a snapshot or SS), of the teacher's interactions with the class. These observations are used to measure the following variables: (FMI) who initiated the interaction, to whom the interaction was directed, what the interaction contains, and how the interaction is framed, and (SS) what activity the teacher was involved in, the materials being used, and with whom the teacher is working. The variables as listed in the authors' checklist, all tend to follow the definitions of an effective teacher.

Micro measures focuses on the interactions between individual students and the teacher. The micro measure uses a 14-second "look" and a 6-second "record" to identify the following variables: interactive and noninteractive that are either academic or nonacademic. These interactions are in terms of the teacher's questioning, providing of information, guidance, reinforcements, corrections, and negative responses. In using this model in assessing teachers, the authors found that the more effective teachers (as defined by the research) tended to (a) ask individual students academic questions at a rate more than twice that of the less effective teachers, (b) deliver almost twice the academic reinforcement to individual students than the less effective teachers, and (c) maintain classrooms where the students were less likely to be using academic materials than in the classrooms of the less effective teachers. The effective teachers spent more time in interactive instruction, less time in organizing and managing, more time on-task, and had students who were more involved in interactive instruction. All of this mirrors the definitions of an effective teacher and current trends in curriculum reforms.

The implications of this research would necessitate evaluators to focus on the micro and macro interactions when conducting classroom observations. This is counter to the idea of just

"looking" as suggested by Dunkleberger (1982). The interactions occurring within the classroom are far to complex to place simple observations as a means of assessing the quality of learning and teaching that is taking place. While Swank, et al. saw no significant correlation between those teachers who scored high on micro and low on macro (and vice versa), the use of this model as a stand alone method or as an addition to an existing model (STAR, CET, TAP, etc.) is highly recommended.

So how often should a teacher be evaluated and does the number of observations influence the final outcome? The models discussed so far all tend to emphasize multiple observations. These should occur every other year or as needed. Cronin and Capie (1986) suggest that the observation of teachers be determined by the reliability and validity of the scores generated during teacher evaluations. In other words, if there is a discrepancy between scores or ratings of a teacher based on observations, what factors are influencing these variations - the teacher's actual performance or factors involving time of observation and/or the evaluator? The authors found that observations on separate days were better predictors of performance than multiple observations on a single day. In addition, variation on scores or ratings from day to day were greater than the variation from observer to observer. The indications from this study suggest that multiple assessors observing on different days may offer the best evaluation of a teacher's performance.

Stodolsky (1984) addresses the issue of teacher observation and reliability from a different perspective. Stodolsky finds fault in classroom observations that reflect chosen subjects by the teacher or the evaluator to be used for evaluations. Each teacher is seen as having strengths or weaknesses within a certain subject or subjects, and that when evaluations are based on these observations, assessments of teaching performance, while reliable, are not really valid. Additionally the teaching of specific subjects necessitates certain teaching behaviors that may not be in evidence in other curriculum subjects being observed (as suggested by TAP). If a teacher is evaluated during a lesson that does not require or elicit a certain behavior (in the teacher and/or the students), then the teacher may receive an unfair evaluation. If multiple evaluations are based on the observation of the same curriculum subject the teachers should receive constant scores - either

high, average, or low. If we consider Cronin and Capie's assumption on extraneous factors influencing learning, then scores may vary from observation to observation. If the same curriculum subject was evaluated each time, other factors may also be involved. If the scores vary from observation to observation, the use of different curriculum subjects may be the factor for variations in scores, in addition to extraneous factors.

The need then, is for a generalizability of teacher assessments by requiring assessing teachers across multiple subjects over periods of time by means of student responses. Student responses were seen as their involvement in the lesson or their ability to be on-task. Stodolsky suggests that by measuring student responses to various subject lessons, inconsistencies in teacher behaviors will be more readily apparent, providing a more consistent picture of a teacher's performance. Stodolsky offers substantial evidence to support this model as a viable measure of teacher performance. Clearly the use of "showcase" lessons cannot provide a valid assessment of teachers. The need now is for teacher assessments to be based on various subject lessons, as measured by student involvement or on-task behavior, during frequent visits.

Natriello (1984) suggests that as the number of observations or evaluations of teachers increase, so will the effectiveness of the teacher. This will occur, it is assumed, through an increase in the reflection into the actions of and by the individual (self-analysis). Natriello equates the effectiveness of a teacher as having "leverage" over what happens in the classroom and the "product" or learning taking place. This "leverage" can be defined as control over the totality of the profession. Natriello suggests research that points to the belief that most teachers feel evaluations may be power moves by the administration, dictating to the teacher what and how to teach. As evaluations increase in number or frequency, the teacher's feeling of "leverage" will decrease creating a teacher who feels that they must appease and perform to the expectations of the administration. Newton and Braithwaite (1988) also suggest that teachers perceive evaluations as a means of bureaucratic control by administrators, when asked for their perceptions on teacher evaluations.

Results from Natriello's study were inconclusive, but showed a positive relationship between the frequency of observation or evaluation and teacher's "leverage" in the classroom. The increase in "leverage" was also related to an increase in teacher effectiveness. The results from this study may relate to the fact that the majority of teachers enjoy classroom visits by administrators, allowing the teacher to demonstrate their abilities as an educator. There are, no doubt, teachers who fear visits by administrators or guests for various reasons. Lack of self-worth, confidence, and teaching ability are all contributors to this loathing of classroom visitations. While this may bring up the issues as suggested by Stodolsky, frequent visits that involve different times are preferable for assessment purposes.

Newton and Brathwaite (1988) found the majority of teachers responding "as often as is necessary, with no set limits" to a survey on the number of observations preferred by teachers. This offers some support for the suggestions of Natriello, in that teachers may actually prefer frequent visits to none at all. This author would rather see an administrator more frequently, than rarely or not at all. Knowing that frequent visits may be the norm, this author sees teachers striving to always trying to present their best.

Assessors, though can bring biases with them when they enter a classroom (Redinger, 1988). These biases may be of a positive nature, in that the evaluator sees the evaluatee as an excellent teacher based on past achievements or common philosophical beliefs, or even due to being a close friend. Biases of a negative nature exist, in that the evaluatee may be perceived as being an ineffective teacher, or is disliked for personal reasons. The first perception dictates that the teacher may be able to do no wrong, while with the latter, the teacher can do no right. Both perceptions will cause an unfair or invalid evaluation, with neither being less harmful than the other. It is possible avoid rater biases with the use of multiple assessors or clearly defined criteria that involves little personal interpretation, but this may be impractical or simply not available due to the evaluation method used at the school site.

Ligion and Ellis (1986) found that rater or evaluator biases could be eliminated through statistical methods. To do this, the authors ranked teachers from highest to lowest as based on

their annual performance evaluations for three years. Each teacher's raw score average was converted to a z-score within all evaluations from the years of the study. The teacher's z-score was then compared with their evaluation or rating. The results clearly identified that biases were in effect and that the use of the statistical method could adjust for rater bias.

Additionally, Ligion and Ellis found that student achievement scores related positively with teachers who had high z-scores, but below average raw scores. Where z-scores and raw scores disagreed, the z-scores categorized teachers as above average who actually had positive achievement discrepancies even though their raw score ratings were below average. Also, those teachers who had above average raw score ratings but below average z-scores, had negative student achievement discrepancies. This is a fairly radical solution and, as the authors found out, very political. The identification of rater bias and evidence that raw scores or the teacher's initial ratings where not valid or reliable, did and should cause great concern for teacher evaluations. While the practical implications of this study may be difficult to realize, the need to curtail rater bias in teacher performance ratings is paramount. The use of multiple assessors also would limit the effect of bias.

It should be noted that while the majority of evaluation models use some form of rating scale, these scales must not be taken literally. Striefer (1987) suggests, that the use of rating scales are only acceptable if they are based on a descriptor system of achievement. Rating scales for the evaluation of teachers typically include the following: outstanding, good, average, fair, and poor. These scales are open to interpretation and require the evaluator to arbitrably decide how to place a teacher's observed behavior. If the teacher is observed as being marginal, in that they do not fall clearly into one rating, then which direction should the teacher be placed? This author sees another area of even greater concern. While already having touched on this in the introduction, this deals with the setting of low expectations. Teachers should never fall into the rating of poor or fair, and fortunately few ever do. These teachers should be terminated if there can be no quick remediation of the problem. Those who do receive ratings of average or good, are actually no better than the teachers who are falling into the fair to poor categories. These teachers are simply

maintaining the status-quo and are offering their students an average education, with results being average students. Rating scales must emphasize that scores below outstanding or exceeds the district's expectations are signs of inadequacy as a professional educator. Higher standards for teachers are a must if school reform is to take place. The fear is that if districts assume this position, than they will only water down the evaluation process, so that more teachers will receive the higher ratings. The need then is for a set of national standards for teaching, thereby avoiding this pratfall of the system.

The rating scales as offered by the STAR system for a teacher being either "acceptable" or "unacceptable" are preferable to those of the CET where teachers are seen as either exceeding district expectations, meeting the districts expectations, needing improvement, or being unsatisfactory. The STAR system is very absolute - either a teacher is doing the job or they aren't. Maybe a rating of "acceptable" should refer to those who are effective in the classroom, "unacceptable, but meets the minimum requirements" for those are are maintaining the status-quo, and with "unacceptable" for those who are ineffective. The CET allows for too much leeway in rating a teacher, and relates to the weaknesses as suggested earlier by this author and those of Striefer. We can no longer allow the profession to accept mediocrity in the classroom, and the ratings of the CET do just that.

An option to the traditional rating scales is offered by Striefer (1987). Striefer suggests that these scales be changed to a system based on the evaluator deciding whether (a) the skill was satisfactorily demonstrated, (b) the skill was not satisfactorily demonstrated but should have been, or (c) there was no opportunity to demonstrate the skill. The system reflects the idea of teachers either performing the job effectively, performing ineffectively, or not doing the job at all. A rating of "there was no opportunity to demonstrate the skill" does not imply that the teacher was not performing effectively, just that the lesson did not lend itself to the display of the targeted behavior. This relates back to the concerns of TAP and Stodolsky to assess each lesson as a seperate entity that is unique unto itself. The requirement of such a system, will be for trained evaluators who can

reliable assess the desired skills and do so objectively. In addition evaluation models will need to be modified if this rating scale can not be easily inserted into the existing system.

From the preceding discussions of evaluation models and approaches, there appears to be no one clear method that can best assess a teacher completely. Striefer (1987) sees a need for an evaluation perspective that entails a framework in which all of the important teaching criteria and instructional models can be successfully operationalized. Teacher evaluation will cover a wide range of behavior from sound pedagogy to coming to work on time, yet the system must retain flexibility to allow for the incorporation of the many outstanding instructional models that become available. Evaluation methods must be both summative and formative to effect a change in the teacher's future performance.

An effective method of evaluating teachers should include the following:

•Evaluation criteria must reflect the current research into what constitutes an effective teacher, that is shared and agreed upon by all parties involved in the evaluation process.

•The evaluation process needs to include, but is not limited to, the following methods of assessment as suggested by the research; classroom observations (micro and macro), student evaluations, measures of student involvement and achievement, and teacher portfolios.

•The method should include multiple evaluators and multiple observations of various subjects over significant periods of time. These observations should be formal and informal in nature.

•The evaluation process must follow clearly defined policies and procedures to insure fairness.

•Rating scales should reflect higher standards agreed upon by all parties involved in the evaluation process.

•Individual results need to be shared through open channels of communication between the teacher and evaluators to affect positively professional attitudes and teaching effectiveness.

•School-wide results need to be used for generating staff developments with the intent of increasing student achievement.

WHAT TO DO WITH EVALUATION RESULTS

If we were to follow a suggestion by Blecke (1982), that if teachers were true professionals, they would be self-analytical and would recognize the areas in which they need the most improvement, and they would also be vitally interested in their own improvement for the sake of doing a better job, the intent of this paper would be a moot point. While there are no doubt teachers who fit the description by Blecke, fortunately the vast majority do not. These are teachers who do have a vested interest in the educational community and are vitally interested in the improvement of their profession. Teachers do assess themselves and do strive for improvement of their craft. Unfortunately, the educational community does not traditionally consider teachers as true professionals (sometimes deservedly so) and imposes its own method of analysis. The decision making authority tends to lie outside of the classroom as to evaluations, curriculum or instructional focus, staff development, and other extrinsic factors that can determine student achievement.

In determining the use of evaluation results, the research by Newton and Braithwaite (1988), Redinger (1988), Pembroke and Goedert (1982), Natriello (1984), and Bingham, Heywood, and White (1991) all lead to furthering the teacher as an effective educator. The primary purpose of teacher evaluation must be to improve the educational system by means of identifying effective teachers, what they are doing, and how this can be taught to those teachers who are evaluated as being less than effective. Additionally those teachers who are less than effective must take the opportunity to self-assess their weaknesses and seek the needed instruction to improve their craft. This is difficult for most teachers (and in general, for most people) to do, as no one wants to admit their shortfalls as an educator.

Results from teacher evaluations must be used for the following:

- To be shared with the individual to reinforce perceptions of self-worth.
- To be shared with the individual in addressing teacher shortfalls or inadequacies that would hinder performance in the classroom.

•To identify effective and ineffective teaching strategies, materials, and/or curriculums.

•To develop staff developments or in-services to increase teacher effectiveness and student achievement.

•To identify teachers for advancement and/or merit awards.

CONCLUSIONS AND RECOMMENDATIONS

The area of teacher evaluation is a complex issue, with many diverse factors influencing the educational outcomes in the classrooms today. While society cannot lay total blame on the educators for the perceived failure of the school system, we must as a profession accept partial responsibility. The educational system today, has a greater responsibility than has ever been seen in the last century. With a society and world in constant flux, heading into the 21st Century at breakneck speed, and while placing greater and greater burdens on the educational system, will require more effective teachers and schools than ever before.

Before states and school districts evaluate teachers they must evaluate the process at hand. If the evaluation process cannot meet the demands of the educational profession and those of society, it is of little use in fulfilling the needs of our students. Districts and states must take a step back and assess their methods for evaluating teachers before we are expected to reap the benefits of today's school reforms. Evaluation methods, as used by many states and districts, have not been changed in the last 20 years. The suggestions and research outcomes as discussd in this paper, should allow states and districts to move teacher evaluation into the 21st Century.

BIBLIOGRAPHY

Alexandrov, Diane. "Teacher Evaluation in an Era of School Reform." Evaluative Report. 1989.

Bingham, Richard, and Others. "Evaluating Schools and Teachers Based on Student Performance." Evaluation Review, Apr. 1991. v15 n2. 191-218

Blecke, A. E. "Encouraging Teacher Support for teacher Evaluations." NASSP December 1982. 16-18

Buttram, Joan, and Bruce Wilson. "Promising Trends in Teacher Evaluation." Educational Leadership, April 1987. 4-6

Capie, William, and Linda Cronin. "How Many Teacher Performance Criteria Should There Be?" Paper presented at the Annual Meeting of the American Educational Research Association. Apr. 1986.

Collins, Angelo. "Transforming the Assessment of Teachers: Notes on a Theory of Assessment for the 21st Century." Paper presented to the Annual Meeting of the National Catholic Education Association, Apr. 1990.

Cronin, Linda, and William Capie. "The Influence of Daily Variation in Teacher Performance on the Reliability and Validity of Assessment Data. Paper presented at the Annual Meeting of the American Educational Research Association. Apr. 1986.

Davey, Bruce. "Evaluating Teacher Competence Through the Use of Performance Assessment Tasks: An Overview." Journal of Personnel Evaluation in Education. n5, 1991. 121-132

Dunkleberger, Gary. "Classroom Observations-What Should Principals Look For?" NASSP, December 1982. 9-15

Ernest, Patricia. "Educational Technology and Teacher Competence: Identification and Preservice Assessment. Paper presented at the Annual Meeting of the Association for Educational Communications and Technology. May 1982.

Herrmann, Beth. "Effective Teacher Evaluation: A Quantitative and Qualitative Process." NASSP, Dec. 1987. 23-30

Hill, Flo, and Others. "Classroom Assessments of 6000 Teachers: What Do the Results Show about the Effectiveness of Teaching and Learning?" Paper presented at the Annual Meeting of the American Educational Research Association. Apr. 1991.

Ligon, Glynn, and John Ellis. "Adjusting for Rater Bias in Teacher Evaluations: Political and Technical Realities." Paper presented at the Annual Meeting of the American Educational Research Association. Apr. 1986.

Million, Steven. "Demystifying Teacher Evaluation: The Multiple-Strategies Model Used as an Assessment Device." Paper presented at the Annual Meeting of the National Council of States on Inservice Education. San Diego, CA. Nov. 1987.

Natriello, Gary. "Teachers' Perceptions of the Frequency of Evaluation and Assessments of Their Effort and Effectiveness." American Educational Research Journal. Fall 1984, v21 n3. 579-595

Newton, E. H., and W. E. Braithwaite. "Teacher Perspectives on the Evaluation of Teachers." Educational Studies 1988. v14 n3. 275-288

Pembroke, Eileen, and Edmund Goedert. "What is the Key to Developing an Effective Teacher Evaluation System?" NASSP, December 1982. 29-37

Redfield, Doris, and James Craig. "Identifying and Documenting Student Outcomes for Use in the Evaluation of Teachers When Standardized Achievement Tests Do Not Apply." Paper presented at the Annual Meeting of the Mid-South Educational Research Association. Nov. 1987.

Redinger, Linda. "Evaluation: A Means of Improving Teacher Performance." Research Report, Apr. 1988.

Savage, John. "Teacher Evaluation Without Classroom Observation." NASSP, Dec. 1982.

Savage, Tom, and Marsha McCord. "The Use of Student Evaluation in the Assessment of Teacher Competence." Paper presented at the Annual Meeting of the American Educational Research Association. Apr. 1986.

"Schools for the 70's and Beyond: A Call to Action." A Staff Report, published by the NEA, 1971.

Stodolsky, Susan. "Teacher Evaluation: The Limits of Looking." Educational Researcher, Nov. 1984. v13 n9. 11-18

Striefer, Phillip. "Teacher Evaluation Systems: A Review of Critical Issues and the Current State of the Art." Regional Laboratory for Educational Improvement of the Northeast and Islands. Aug. 1987.

Swank, Paul, and Others. "Sensitivity of Classroom Observation Systems: Measuring Teacher Effectiveness. Journal of Experimental Education, Winter 1989. v57 n2. 171-86

Vogt. Craig. "Developing a Teacher Evaluation System." Spectrum Winter 1984, v2 n1. 41-46

White, Kinnard, and others. "Assessing Teacher Performance Using an Observation Instrument Based on Research Findings." NASSP, Mar. 1987. 89-95